BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 4: e59074
doi: 10.3897/biss.4.59074

OPEN ACCESS

Conference Abstract

The Automated Taxonomic Concept Reasoner

Atriya Sen‡, Nico Franz‡, Beckett W. Sterner‡, Nate Upham‡

‡ Arizona State University, Tempe, United States of America

Corresponding author: Atriya Sen (accounts@atriyasen.com)

Received: 28 Sep 2020 | Published: 30 Sep 2020

Citation: Sen A, Franz N, Sterner BW, Upham N (2020) The Automated Taxonomic Concept Reasoner. Biodiversity Information Science and Standards 4: e59074. https://doi.org/10.3897/biss.4.59074

Abstract

We present a visual and interactive taxonomic Artificial Intelligence (AI) tool, the Automated Taxonomic Concept Reasoner (ATCR), whose graphical web interface is under development and will also become available via an Application Programming Interface (API). The tool employs automated reasoning (Beeson 2014) to align multiple taxonomies visually, in a web browser, using user or expert-provided taxonomic articulations, i.e. "Region Connection Calculus (RCC-5) relationships between taxonomic concepts, provided in a specific logical language (Fig. 1). It does this by representing the problem of taxonomic alignment under these constraints in terms of logical inference, while performing these inferences computationally and leveraging the powerful Microsoft Z3 Satisfiability Modulo Theory (SMT) solver (de Moura and Bjørner 2008). This tool represents further development of utilities for the taxonomic concept approach, which fundamentally addresses the challenge of robust biodiversity data aggregation in light of multiple conflicting sources (and source classifications) from which primary biodiversity data almost invariably originate. The approach has proven superior to aggregation, based just on the syntax and semantics provided by the Darwin Core standard Franz and Sterner 2018).

Fig. 1 provides an artificial example of such an alignment. Two taxonomies, A and B, are shown. There are five taxonomic concepts, A.One, A.Two, A.Three, B.One and B.Two. A.Two and A.Three are sub-concepts (children) of A.One, and B.Two is a sub-concept (child) of B.One. These are represented by the direction of the grey arrows. The undirected mustard-coloured lines represent relationships, i.e., the articulations referred to in the previous paragraph. These may be of five kinds: congruent (==), includes (<) and included in (>), overlap (><), and disjointness. These five relationships are known in the AI literature

© Sen A et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

as the Region Connection Calculus-5 (RCC-5) (Randell et al. 1992, Bennett 1994, Bennett 1994), and taken exclusively and in conjunction with each other, have certain desirable properties with respect to the representation of spatial relationships. The provided relationship (i.e. the articulation) may also be an arbitrary disjunction of these five fundamental kinds, thus allowing for representation of some degree of logical uncertainty. Then, and under three assumptions that:

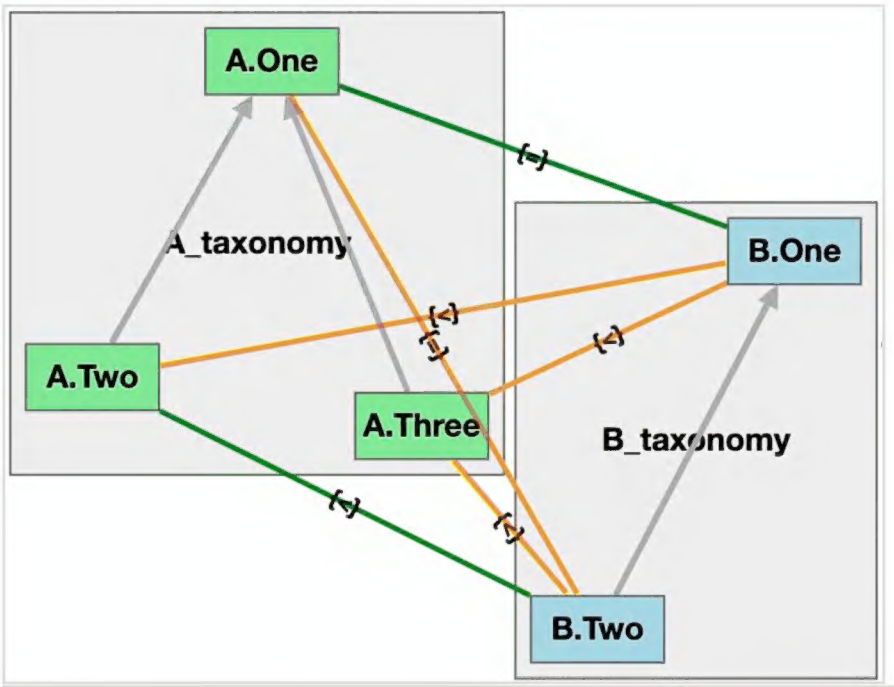

Figure 1.

An example of the Automated Taxonomic Concept Reasoner (ATCR) web interface and functionality. Shown are two input taxonomies (A, B) with three and two entailed concept regions, respectively. Each of these stands for a taxonomic concept as recognized and delimited by the respective source. The grey arrows symbolize given parent-child relationships within each input taxonomy. Green lines show user-specified input RCC-5 articulations. Mustard-coloured lines show logically contingent, reasoner-inferred articulations. The example is logically consistent; if it were not, then no mustard-coloured lines would be visualized.

1. "sibling" concepts are disjoint in their instances,
2. all instances of a parent concept are instances of at least one of its child concepts, and
3. every concept has at least one instance - the SMT-based automated reasoner is able to deduce the relationships represented by the undirected green lines. It is also able to deduce disjunctive relationships where these are logically implied.

ATCR is related to Euler/X (Franz et al. 2015), an existing tool for the same kinds of taxonomic alignment problems, which was used, for example, to obtain an alignment of two influential primate classifications (Franz et al. 2016). It differs from Euler/X in that it employs a different logical encoding that enables more efficient and more informative computational reasoning, and also in that it provides a graphical web interface, which Euler/X does not.

Keywords

automated reasoning, artificial intelligence, taxonomic intelligence, computational systematics, bioinformatics, biodiversity informatics

Presenting author

Atriya Sen

Presented at

TDWG 2020

References

- Beeson M (2014) Larry Wos, Ross Overbeek, Ewing Lusk, and Jim Boyle. Automated reasoning. Introduction and applications. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1984, xiv + 482 pp. Journal of Symbolic Logic 51 (2): 464-465. https://doi.org/10.1017/s0022481200031340
- Bennett B (1994) Spatial Reasoning with Propositional Logics. Principles of Knowledge Representation and Reasoning 51-62. https://doi.org/10.1016/b978-1-4832-1452-8.50102-0
- de Moura L, Bjørner N (2008) Z3: An Efficient SMT Solver. Tools and Algorithms for the Construction and Analysis of Systems 337-340. https://doi.org/10.1007/978-3-540-78800-3_24
- Franz N, Chen M, Yu S, Kianmajd P, Bowers S, Ludäscher B (2015) Reasoning over Taxonomic Change: Exploring Alignments for the Perelleschus Use Case. PLOS ONE 10 (2). https://doi.org/10.1371/journal.pone.0118247
- Franz NM, Pier NM, Reeder DM, Chen M, Yu S, Kianmajd P, Bowers S, Ludäscher B (2016) Two Influential Primate Classifications Logically Aligned. Systematic biology 65 (4): 561-82. https://doi.org/10.1093/sysbio/syw023
- Franz NM, Sterner BW (2018) To increase trust, change the social design behind aggregated biodiversity data. Database 2018 https://doi.org/10.1093/database/bax100
- Randell DA, Cui Z, Cohn AG (1992) A spatial logic based on regions and connection. In: Nebel B, Swartout W, Rich C (Eds) Proceedings of the 3rd International Conference on Knowledge Representation and Reasoning. Morgan Kaufmann, Los Altos, 165-176 pp.